UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/600,475	06/20/2003	Chris J.C. Burges	MCS-018-03	6335

7590 02/07/2008

LYON & HARR, L.L.P
Suite 800
300 Esplanade Drive
Oxnard, CA 93036-1274

EXAMINER	
SIEDLER, DOROTHY S	
ART UNIT	**PAPER NUMBER**
2626	

MAIL DATE	DELIVERY MODE
02/07/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)
	10/600,475	BURGES ET AL.
	Examiner	Art Unit
	Dorothy Sarah Siedler	2626

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on 22 October 2007.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) 1,3,5-16,19,20 and 22-27 and 29 is/are pending in the application.

4a) Of the above claim(s) 2,4,17,18,21 and 28 is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) 1,3,5-16,19,20,22-27 and 29 is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☐ The drawing(s) filed on _____ is/are: a)☐ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Response to Arguments

Applicant's arguments with respect to ***Sturim***, filed October 22, 2007 have been fully considered but they are not persuasive.

Applicant argues that, "Sturim et al merely teach using a Gaussian Mixture Model (GMM) as the anchor model.", however the examiner respectfully disagrees. ***Sturim*** specifically states, "The basic concept of anchor modeling is the representation of a target speech utterance with information gained from a set of models pre-trained from a defined set of talkers. In theory, the models could consist of virtually any method of speech representation." (section 2, first paragraph). ***Sturim*** does disclose previous work on anchor modeling using Hidden Markov Models, and the current work using GMM-UBM, however, based on the statement recited above, ***Sturim*** clearly suggests the use of any method of speech representation in anchor modeling.

The examiner notes applicant's response to the 35 U.S.C. 112 second paragraph rejection of claim 18 now incorporated into claim 14. However, the amendment was not sufficient to clarify the ambiguity, therefore the rejection is maintained.

The remainder of Applicant's arguments with respect to claims 1-4, 6-10,12-14,17-19 and 20-28 have been considered but are moot in view of the new ground(s) of rejection.

Requirement for Information – 37 CFR §1.105

Applicant and the assignee of this application are required under 37 CFR 1.105 to provide the following information that the examiner has determined is reasonably necessary to the examination of this application.

In response to this requirement, please provide a copy of each of the following items of art referred to in the specification: chapter six of "Pattern Classification" 2nd edition, by R.O. Duda, P.E. Hart, and D.G. Stork as well as chapter six of "Neural Networks for Pattern Recognition" by C.M. Bishop.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 14 and 24 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 14 recites, "..the plurality of anchor models comprising discriminatively-trained classifiers of a convolutional neural network that were previously trained using a training technique that included non-linear terms", however the phrase "non-linear terms" is ambiguous. It is unclear whether the phrase "non-linear terms" refers to the input training data, which would render the claim nonsensical, or to the non-linear output function used in most neural network structures. Based on the specification, the

examiner interprets the phrase "non-linear terms" as “final nonlinearity process". This interpretation is used throughout the remainder of this office action.

Claim 24 recites similar limitations, and is therefore rejected fro similar reasons.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1, 3, 6-10, 12-14,16, 20, and 22-27 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** (“Speaker Indexing in Large Audio Databases Using Anchor Models” 2001) in view of ***Waibel*** (“Phoneme Recognition Using Time-Delay Neural Networks” IEEE 1989).

As per claim 1, ***Sturim*** discloses a method for processing audio data, comprising:

applying the plurality of anchor models to the audio data (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers*);

mapping the output of the plurality of anchor models into frame tags and producing the frame tags (section 2. Anchor Models, *speaker characterization vectors*

are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space).

Sturim does not disclose using discriminatively-trained classifiers that are time-delay neural network (TDNN) classifiers to produce a plurality of anchor model outputs. However, ***Sturim*** does disclose that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Waibel*** discloses a speech processing system that uses a TDNN for the speech representation. (Abstract). ***Sturim*** and ***Waibel*** both disclose system for improved speech processing, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a TDNN as an anchor model in ***Sturim***, since the time-delay structure enables the system to discover the temporal relationship among acoustic features independent of the position in time, as indicated in ***Waibel*** (Abstract).

As per claim 20, ***Sturim*** discloses a method for processing audio data containing a plurality of speakers, comprising:

applying a plurality of anchor models to the audio data (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers*);

mapping an output of the anchor models into frame tags (section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*); and

a training set containing a set of training speakers, and wherein the plurality of speakers is not in the set of training speakers (section 1. Introduction, *speakers of the target utterance are not members of the training* set).

Sturim does not explicitly state constructing a list of start and stop times for each of the plurality of speakers based on the frame tags, nor using discriminatively-trained classifiers that are time-delay neural network TDNN) classifiers to produce a plurality of anchor model outputs. However, ***Sturim*** does disclose a system that can be used to retrieve messages from an archive (section 4. Speaker Indexing). In order to retrieve the messages, the system must know where the messages start and stop, and therefore must determine start and stop times for each speaker. ***Sturim*** also discloses that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Waibel*** discloses a speech processing system that uses a TDNN for the speech representation. (Abstract). ***Sturim*** and ***Waibel*** both disclose system for improved speech processing, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to construct a list of start and stop times in ***Sturim***, since start and stop

times can be used to reliably retrieve and playback saved messages corresponding to a specific speaker.

Therefore it would also have been obvious to one of ordinary skill in the art at the time of the invention to have a TDNN as an anchor model in ***Sturim***, since the time-delay structure enables the system to discover the temporal relationship among acoustic features independent of the position in time, as indicated in ***Waibel*** (Abstract).

As per claim 3, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claim 2, and ***Sturim*** further comprising training the TDNN classifier on data separate from audio data available in a use phase (section 1. Introduction, *speakers of the target utterance are not members of the training* set).

As per claim 6, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claim 1, and ***Waibel*** further discloses pre-processing the audio data to generate input feature vectors for the discriminatively-trained classifier (section II A, *melscale spectral coefficients are derived from the input speech, then input to the network*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the audio data to generate input feature vectors in ***Sturim***, since it would provide a reliable set of feature vectors, which can be easily applied to the classifier for further processing.

As per claims 7, 8 and 22, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claims 1 and 20 and ***Sturim*** further discloses normalizing a feature vector output of the discriminatively-trained classifier (section 2. Anchor Models, second paragraph, *each anchor model yields a likelihood score, where the combination of scores are used to form a N-dimensional characterization vector*, and the fifth paragraph to the sixth paragraph, *a comparison is done between normalized data and non-normalized output data, therefore normalization must have been done*). ***Sturim*** does not explicitly state wherein the normalized feature vectors are vectors of unit length. However Official notice is taken that it is old and well known to normalize a vector to a vector of unit length. During vector processing, feature vectors are often normalized to a unit vector for simplicity of computation.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to normalize a vector output of the classifier to a unit vector in ***Sturim***, since it would produce a simplified feature vector, enabling simplified processing which then reserves computational resources.

As per claim 9, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claim 1, however ***Sturim*** does not explicitly disclose accepting a plurality of input feature vectors corresponding to audio features contained in the audio data, and applying the discriminatively-trained classifier to the plurality of input feature vectors to

produce a plurality of anchor model outputs. However, ***Sturim*** does disclose applying input data to a trained anchor models to produce anchor model outputs (section 2. Anchor Models). In addition, ***Waibel*** discloses accepting a plurality of feature vectors corresponding to audio features contained in the audio data (section II A, *melscale spectral coefficients are derived from the input speech, then input to the network*), and applying the discriminatively-trained classifier to the plurality of input feature vectors to produce a plurality of model outputs (Abstract, *a TDNN is used for speech processing*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to accept a plurality of input feature vectors corresponding to audio features contained in the audio data, and apply the discriminatively-trained classifier to the plurality of input feature vectors to produce a plurality of anchor model outputs in ***Sturim***, since it would provide a reliable set of feature vectors which can be easily applied to a TDNN, where the time-delay structure enables the system to discover the temporal relationship among acoustic features independent of the position in time, as indicated in ***Waibel*** (Abstract).

As per claim 10, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claim 1, and ***Sturim*** further discloses wherein the mapping comprises clustering anchor model outputs from the discriminatively-trained classifier into separate clusters using a clustering technique, and associating a frame tag to each separate cluster (section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A*

determination of the speaker is made based on the location of the vector within speaker space).

As per claim 12, ***Sturim*** in view of ***Waibel*** disclose the method as set forth in claim 1, and ***Sturim*** further discloses training the discriminatively-trained classifier using a speaker training set containing a plurality of known speakers (section 1. Introduction, *anchor models are derived from a set of predetermined speakers*). ***Sturim*** does not explicitly disclose pre-processing the speaker training set and the audio data in the same manner to provide a consistent input to the discriminatively trained classifier. However, ***Waibel*** discloses pre-processing of audio data (section II A, *melscale spectral coefficients are derived from the input speech, then input to the network*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the speaker training set and the audio data in the same manner in ***Sturim***, since it would provide reliable data input to the classifier, which would then provide a reliable and useful result.

As per claims 13 and 23, neither ***Sturim*** in view of ***Waibel*** explicitly disclose computer-readable medium having computer-executable instructions for performing the method recited in claims 1 and 20. However, the method of ***Sturim*** requires considerable computation and processing, and modern computer systems can perform the same computations considerably faster, and with higher accuracy, than any human

would. In addition, ***Waibel*** states that the disclosed system was implemented using C and Fortran (page 331, second column), both common programming languages used to execute computer readable instructions.

Therefore it would have been obvious to perform the method of claim 1 on a computer-readable medium in ***Sturim***, since a computer would enable faster processing, saving time and assuring accuracy.

As per claim 14, ***Sturim*** discloses a computer-implemented process for processing audio data, comprising:

applying a plurality of anchor models to the audio data (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers*);

normalizing the modified feature vector output to generate normalized anchor model output (section 2. Anchor Models, second paragraph, *each anchor model yields a likelihood score, where the combination of scores are used to form a N-dimensional characterization vector*, and the fifth paragraph to the sixth paragraph, *a comparison is done between normalized data and non-normalized output data, therefore normalization must have been done*);

mapping the normalized anchor model output into frame tags and producing frame tags (section 2. Anchor Models, *speaker characterization vectors are mapped*

onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space).

Sturim does not disclose the plurality of anchor models comprising discriminatively-trained classifiers of a convolutional neural network that were previously trained using a training technique that included non-linear processing step. However, ***Sturim*** does disclose that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Waibel*** discloses a speech processing system that uses a TDNN for speech representation (Abstract), where a TDNN is a type of convolutional neural network. The TDNN of ***Waibel*** is trained using the sigmoid function as the non-linear output function (section II A, first paragraph and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use a discriminatively-trained classifiers of a convolutional neural network that was previously trained using a training technique that included non-linear processing step in ***Sturim,*** since the time-delay structure enables the system to discover the temporal relationship among acoustic features independent of the position in time, as indicated in ***Waibel*** (Abstract).

Neither ***Sturim*** nor ***Waibel*** disclose obtaining a preliminary output of the plurality of anchor models from the convolutional neural network before final nonlinearity process is applied to generate a modified feature vector output. However, ***Sturim*** disclose the use of anchor models, where anchor models are trained classifiers and the output is

input into another machine-learning algorithm, such as a clustering algorithm. In addition, neural networks can be designed for pattern recognition when the number of pattern classes is known, or used prior to clustering when the number of pattern classes is unknown. During pattern classification with known pattern classes a sigmoid, or other non-linear function, is used as a final output function. This enables the system to provide a specific result of the classification, indicating the most likely class; However, when the number of classes in unknown the desired neural network output is not a specific class value. Instead, each output of the neural network is clustered and analyzed to determine the class values for the unknown input. ***Sturim*** discloses a system for speaker indexing using anchor models designed with a classification and clustering step. Previously trained speaker anchor models are used to project target vectors onto a speaker space defined by the anchor models. Speaker detection is performed by analyzing the vectors within the speaker space.

Therefore it would also have been obvious to one of ordinary skill in the art at the time of the invention to obtain a preliminary output of the plurality of anchor model from the convolutional neural network before the final nonlinearity process is applied to generate a modified feature vector output in ***Sturim***, since one of ordinary skill in the art has good reason to pursue the options within his or her technical grasp in order to achieve the predictable result of determining optimum feature vectors to map into the speaker space, thus improving the accuracy of the speaker indexing.

As per claim 16, ***Sturim*** in view of ***Waibel*** disclose the system of claim 14, and ***Waibel*** further discloses wherein the training technique employs a mean-square error metric (section II B, first paragraph). ***Waibel*** also discloses that there are many learning techniques for the optimization of neural networks, including mean-squared error.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use the mean-square error during training in ***Sturim***, since one of ordinary skill in the art at the time of the invention has good reason to pursue the options within his to her technical grasp.

As per claim 19, this claim recites limitations similar to claim 8, and is therefore rejected for similar reasons.

As per claim 24, ***Sturim*** disclose a computer-readable medium having computer-executable instructions for processing audio data, comprising:

training anchor models to be used to produce anchor models outputs, and (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers* and section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*).

normalizing the modified plurality of anchor model output to generate normalized anchor model outputs (section 2. Anchor Models, second paragraph, *each anchor model yields a likelihood score, where the combination of scores are used to form a N-dimensional characterization vector*, and the fifth paragraph to the sixth paragraph, *a comparison is done between normalized data and non-normalized output data, therefore normalization must have been done*);

clustering anchor model outputs into frame tags of speakers (Section 1. Introduction, *a target utterance is characterized using anchor models derived from a predetermined set of speakers* and section 2. Anchor Models, *speaker characterization vectors are mapped onto a speaker space. A determination of the speaker is made based on the location of the vector within speaker space*).

Sturim does not disclose training a discriminatively-trained classifier that is a time-delay neural network (TDNN) in a discriminative manner on a convolutional neural network using a training technique that includes a non-linear processing step such that the training occurs during a training phase to generate parameters that can be used at a later time by the TDNN classifier, and using the discriminatively-trained classifiers that are time-delay neural network (TDNN) classifiers to produce a plurality of anchor model outputs. However, ***Sturim*** does disclose that anchor models, previously trained during a training phrase, can consist of any method of speech representation (section 1. Introduction and section 2. Anchor Models, first paragraph). In addition, ***Waibel*** discloses a speech processing system that uses a TDNN for speech representation

(Abstract), where a TDNN is a type of convolutional neural network. The TDNN of ***Waibel*** is trained using the sigmoid function as the non-linear output function (section II A, first paragraph and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to train a discriminatively-trained classifier that is a time-delay neural network (TDNN) in a discriminative manner on a convolutional neural network using a training technique that includes a non-linear processing step such that the training occurs during a training phase to generate parameters that can be used at a later time by the TDNN classifier, using the discriminatively-trained classifiers that are time-delay neural network (TDNN) classifiers to produce a plurality of anchor model outputs in ***Sturim,*** since the time-delay structure enables the system to discover the temporal relationship among acoustic features independent of the position in time, as indicated in ***Waibel*** (Abstract).

Neither ***Sturim*** nor ***Waibel*** disclose obtaining a preliminary output of the plurality of anchor models from the convolutional neural network before final nonlinearity process is applied to generate a modified feature vector output. However, ***Sturim*** disclose the use of anchor models, where anchor models are trained classifiers and the output is input into another machine-learning algorithm, such as a clustering algorithm. In addition, neural networks can be designed for pattern recognition when the number of pattern classes is known, or used prior to clustering when the number of pattern classes is unknown. During pattern classification with known pattern classes a sigmoid, or other

non-linear function, is used as a final output function. This enables the system to provide a specific result of the classification, indicating the most likely class; However, when the number of classes in unknown the desired neural network output is not a specific class value. Instead, each output of the neural network is clustered and analyzed to determine the class values for the unknown input. ***Sturim*** discloses a system for speaker indexing using anchor models designed with a classification and clustering step. Previously trained speaker anchor models are used to project target vectors onto a speaker space defined by the anchor models. Speaker detection is performed by analyzing the vectors within the speaker space.

Therefore it would also have been obvious to one of ordinary skill in the art at the time of the invention to obtain a preliminary output of the plurality of anchor model from the convolutional neural network before the final nonlinearity process is applied to generate normalized anchor model outputs in ***Sturim***, since one of ordinary skill in the art has good reason to pursue the options within his or her technical grasp in order to achieve the predictable result of determining optimum feature vectors to map into the speaker space, thus improving the accuracy of the speaker indexing.

As per claim 25, ***Sturim*** in view of ***Waibel*** disclose the computer-readable medium of claim 24, and ***Waibel*** further discloses pre-processing a speaker training set during the training phase to produce a first set of input feature vectors for the discriminatively-trained classifier (section II A, *melscale spectral coefficients are derived*

from the input speech, then input to the network). However, neither ***Sturim*** nor ***Waibel*** disclose pre-processing a speaker training set during a validation phase to produce a first set of input feature vectors for the discriminatively-trained classifier. However, by applicants own admission (specification page 17, second paragraph) validation sets are old and well known.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the audio data to generate input feature vectors in a training and validation phase in ***Sturim***, since it would provide a reliable set of feature vectors, which can be easily applied to the classifier for further processing.

As per claim 26, ***Sturim*** in view of ***Waibel*** disclose the computer-readable medium of claim 25, however ***Sturim*** does not explicitly disclose pre-processing the audio data during the use phase to produce a second set of input feature vectors for the discriminatively-trained classifier, the pre-processing of the audio data being preformed in the same manner as the pre-processing of the speaker training set. However, ***Waibel*** discloses pre-processing of audio data (However, ***Waibel*** discloses pre-processing of audio data (section II A, *melscale spectral coefficients are derived from the input speech, then input to the network*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to pre-process the speaker training set and the audio data in the same

manner in ***Sturim***, since it would provide reliable data input to the classifier, which would provide a reliable and useful result.

As per claim 27, this claim recites limitations similar to those recited in claim 8, and is therefore rejected for similar reasons.

Claims 5,15 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** in view of ***Waibel*** as applied to claims 4 and 14 above, and further in view of ***Lavagetto*** ("Time-Delay Neural Network for Estimating Lip Movements from Speech Analysis: A useful Tool in Audio-Video Synchronization" IEEE 1997).

Sturim in view of ***Waibel*** disclose the method as set forth in claim 1 and 14, however neither explicitly disclose further training the TDNN classifier using cross entropy. However, by applicant's own admission training using cross entropy is well known (specification page 29). In addition, ***Lavagetto*** discloses that training a time-delay neural network can be done with either cross entropy of mean-square error (page 789-790).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to use cross entropy or mean-square error to train the TDNN in ***Sturim*** and ***Waibel,*** since cross entropy and mean-square error provide figures for validating estimates provided by each network independent from the network structure itself, as indicated in ***Lavagetto*** (page 789,section IV. Learning Criteria for TDNN Training).

Claims 11 and 29 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Sturim*** in view of ***Waibel*** as applied to claims 10 and 25 above, and further in view of ***Liu*** (6,615,170).

Sturim in view of ***Waibel*** disclose the method as set forth in claims 10 and 25, however neither disclose applying temporal sequential smoothing to the frame tag using temporal information associated with the clustered anchor model outputs. ***Liu*** discloses temporal smoothing tagged frames (column 5 line 55- column 5 line 20). ***Liu*** discloses tagging speech frames based on the output of specific model. Adjacent observations are then used to update the vale of a tag for each frame by weighting observations at different times.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to apply temporal sequential smoothing to the frame tags in ***Sturim*** and ***Waibel***, since it enables the incorporation of adjacent frame tags for updating and validating a current frame tag, thus increasing tagging accuracy, as indicated in Liu (column 5 lines 64-65).

Conclusion

This Office action has an attached requirement for information under 37 CFR 1.105. A complete reply to this Office action must include a complete reply to the attached requirement for information. The time period for reply to the attached requirement coincides with the time period for reply to this Office action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER